REMARKS

Claims 1-20 and 23-32 are all the claims pending in the application, and stand rejected on prior art grounds. Claims 1, 23, and 28 are amended herein. Applicants respectfully traverse these rejections based on the following discussion.

I. The Prior Art Rejections

Claims 1-20, and 23-32 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Hattori, et al. (U.S. Patent No. 6,889,223 B2), hereinafter referred to as Hattori, in view of Cruz, et al. ("Measuring Structural Similarity Among Web Documents: Preliminary Results", 1998, Lecture Notes in Computer Science, Volume 1375, page 513), hereinafter referred to as Cruz, in further view of Schuetze, et al. (U.S. Patent No. 6,598,054 B2), hereinafter referred to as Schuetze. Applicants respectfully traverse these rejections based on the following discussion.

Hattori teaches that when a retrieval condition including a first desired word and a first desired component including a value in which the first desired word is included, is inputted, a first detecting device detects second desired components each being similar to the first desired component, an acquiring device acquirers second desired words each being similar to the first desired word, a first retrieving device retrieves first structured documents each including a first component including a value in which one of the first desired word and the second desired words is included, a second retrieving device retrieves second structured documents each including a second component corresponding to one of the first desired component and the second desired components and including or corresponding to the first component.

Cruz teaches a technique for determining structurally similar web pages, and that some

10/629,133   11

structural properties can be identified with semantic properties of the data and provide measure for comparison between HTML documents.

Schuetze teaches a system and method for browsing, retrieving, and recommending information from a collection uses multi-modal features of the documents in the collection, as well as an analysis of users' prior browsing and retrieval behavior. The system and method are premised on various disclosed methods for quantitatively representing documents in a document collection as vectors in multi-dimensional vector spaces, quantitatively determining similarity between documents, and clustering documents according to those similarities. The system and method also rely on methods for quantitatively representing users in a user population, quantitatively determining similarity between users, clustering users according to those similarities, and visually representing clusters of users by analogy to clusters of documents.

However, the claimed invention, as provided in amended independent claims 1, 23, and 28 contain features, which are patentably distinguishable from the prior art references of record. Specifically, claims 1, 23, and 28 provide in part, "…wherein said path representations comprise a path in the tree representation of a document having a capability to include positional information of preceding sibling nodes that have a same label as a given node in a tree…."

These features are neither taught nor suggested in the prior art of record.

The objective of Hattori is "to provide a method and an apparatus which can retrieve structured documents, each document structure of the structured document is equal/similar to that designated in the retrieval condition". It is important to note that in Hattori the notion of equal/similar is based on the values of the components of the structured document. In contrast, the notion of the similarity in Applicants' claimed invention is based on structural properties of

10/629,133   12

the whole document and <u>not</u> on the specific values contained in the document. In contrast to the Applicants' claimed invention, Hattori tries to find similarity between the words given in a retrieval condition with the values contained in a component of a document. Hattori does not consider a document as a whole for computing its similarity with any other document.

Furthermore, the Applicants' claimed invention computes similarity based entirely on the structure of the document and does not use the actual values occurring in documents. Conversely, the notion of similarity taken in Hattori is based on values. Furthermore, Hattori computes the equality/similarity of a word in retrieval condition with the values occurring in document. Conversely, the Applicants' claimed invention computes similarity between two documents. Due to this difference in the objectives, there is a significant difference between the representation of documents and in the process of computing the similarity. In particular, the Applicants' claimed invention represents a document using the paths but does not include the values (that occur as leaf nodes in tree representation of documents).

One of the objectives of the Applicants' claimed invention is to compute the similarity of documents based on their structural similarity only. The Applicants' claimed invention can compare the structural representations of two documents to determine their similarity. For example, the (two) documents to be compared can both be represented in the form of trees (e.g. DOM trees), and then computations are made to determine how similar the trees are with respect to one another. However, as mentioned in the Applicants' specification, this process is computationally very expensive. Accordingly, the Applicants' overcome this because the Applicants' claimed invention captures most of the structural properties contained in the tree representation in the form of a feature vector. In particular, the Applicants' claimed invention

10/629,133   13

captures the structural properties in the form of paths or X-paths. Once the feature vector

representation is computed, a similarity measure can be defined and used to compute the

similarity between two feature vectors.

As described above, in the Applicants' claimed invention a document is represented by

the set of the paths that occur in the document. These paths do not include the values. Further,

Hattori does not use these paths as features for representing documents as do the Applicants. In

the Applicants' claimed invention, a document may be represented using all the "X-paths" that

occur in the document. The Applicants respectfully argue that the suggested similarity between

the Applicants' claimed invention and Hattori (column 8. line 42-57) given on page 7 of the

Office Action is erroneous. Hattori, assigns a unique object id to each node. An X-path contains

a positional index, which represents, for a node (n), the number of previous sibling nodes with

the same label as that of node (n). The notion of "X-paths" is not new but it is used as a query

language for accessing some specific components of a document. Accordingly, the Applicants'

claimed invention uses X-paths in a novel way; to capture the structural properties of the

document in the form of a feature vector.

Representing documents in a vector form is a standard method in text and data mining

community. However, the features that are used in the vector representation determine the type

of similarity that can be computed. Accordingly, the Applicants' claimed invention uses this

concept in a novel way; using paths/X-paths as features and uses them to represent documents

(note that these paths/X-paths are different from the one used in Hattori). In the Applicants case,

two alternative ways of representing a document are provided; one using paths and the other

using X-paths. The similarity measure used when paths are used as features and documents are

10/629,133   14

represented as vectors of these features is a standard similarity measure. In fact, any other known measure of similarity between vectors can also be used. However, for the "X-path" representation, the Applicants provide a novel similarity measure (for example, see claims 14-19). This similarity measure is novel and is not given in Cruz or anywhere else.

Additionally, the Applicants' claimed invention is not obvious in the light of the combination of Hattori, Cruz, and Schuetze for the following reasons: It is not obvious, from any of the literature given for the basis of obviousness, how to represent the structural information of a document in the form of a feature vector. Moreover, the path representation used by the Applicants does not use actual values contained in the document, whereas Hattori's method uses the actual values of the document for computing the similarity. In fact, in the Applicants' case, it is important to ignore the actual values completely. It may be obvious how to use an existing similarity measure to compute similarity between two documents when a document is represented using paths. However, existing measures do not work when the document is represented using X-paths. Thus, in addition to providing a novel and non-obvious method to represent structural information of a document in the forma of a vector, the Applicants also provides a novel similarity measure for the case when an X-Path representation is used. The Applicants refer to pages 7 and 8 of the Applicants' specification for a definition of the term X-Path within the context of the Applicants' claimed invention. Accordingly, the claims must be read in light of the specification and the terms defined therein to which the claimed language is directed. Accordingly, the prior art of record does not teach "wherein said path representations comprise a path in the tree representation of a document having a capability to include positional information of preceding sibling nodes that have a same label as a given node in a tree."

10/629,133   15

Moreover, the Applicants note that all claims are properly supported in the specification and accompanying drawings. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw the rejections.
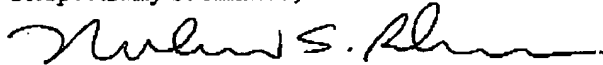
II.   Formal Matters and Conclusion

With respect to the rejections to the claims, the claims have been amended, above, to overcome these rejections. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw the rejections to the claims.

In view of the foregoing, Applicants submit that claims 1-20, and 23-32, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary. Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,

Dated: September 29, 2006

Mohammad S. Rahman
Registration No. 43,029

Gibb I.P. Law Firm, LLC
2568-A Riva Road, Suite 304
Annapolis, MD 21401
Voice: (301) 261-8625
Fax: (301) 261-8825
Customer Number: 29154

10/629,133   16